Renewable and Sustainable Energy Reviews 15 (2011) 2826–2832

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Simulation of correlated wind speeds: A review

Andrés Feijóo [a,*], Daniel Villanueva [a], José Luis Pazos [a], Robert Sobolewski [b]

[a] Departamento de Enxeñería Eléctrica, Universidade de Vigo, ETSEI, Campus de Lagoas-Marcosende, Vigo, Galicia, Spain
[b] Faculty of Electrical Engineering, Bialystok University of Technology, 15-351 Bialystok, Poland

ARTICLE INFO

Article history:
Received 15 December 2010
Accepted 20 February 2011

Keywords:
Wind speed
Wind power
Correlation
Weibull distributions
Rayleigh distributions
Monte Carlo techniques

ABSTRACT

This paper presents the evolution of techniques for simulating correlated wind speeds, over the last decade. The work stems from the problem of obtaining a value that can be defined as the simultaneousness of the production of wind power in electrical networks containing many wind parks. As will be seen, we have steadily extended the research towards the analysis of the correlation of wind speed series at several locations, which is important for assessing the probability of a given wind power being injected in the electrical network.

Contents

1. Introduction 2826
2. Wind speed distributions 2827
3. Simulation of correlated wind speeds according to Pearson parametric correlations 2827
   3.1. The method 2828
   3.2. Error measurement 2829
   3.3. Example 2829
4. An improvement of the method: the use of evolutionary algorithms 2829
   4.1. Description of the method 2829
   4.2. Options to initialize the method 2830
      4.2.1. Fixed initialization 2830
      4.2.2. Optimum initialization 2830
   4.3. Initialization evaluation 2830
5. Simulation of correlated wind speeds according to Spearman rank correlations 2830
   5.1. The method 2830
   5.2. Example 2831
6. Other related works 2831
7. Conclusions 2831
   Acknowledgements 2831
   References 2832

1. Introduction

IT is well known that wind energy has been experiencing important growth in power generation in many countries around the world. Until a few years ago it was mainly in Europe (with Germany, Denmark and Spain in the lead) and the US. Later India joined this trend and, in the last two years, China has come closer and closer to topping the list [1].

When wind energy injection became large scale in the power systems of some of these countries, one major concern for stakeholders, such as the Transmission System Operators (TSO) and the distribution companies, was the idea that wind energy is not easy to control, could not be stored as the primary resource, and was, in some ways, difficult to predict reliably.

* Corresponding author. Tel.: +34 986812055.
E-mail addresses: afeijoo@uvigo.es (A. Feijóo), dvillanueva@uvigo.es (D. Villanueva), pazos@uvigo.es (J.L. Pazos), robsob@pb.edu.pl (R. Sobolewski).

1364-0321/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rser.2011.02.032

Some distribution companies wondered if a concept of simultaneousness could be applied to wind energy production, similar to the way it is defined for the loads. In other words, they wondered if it could be said that a good approximation for wind parks included in power flow analyses could be to consider all them operating at, for example, 40% of their nominal power.

For loads, the idea of groups of them behaving similarly (i.e., operating at a similar percentage of their maximum value) is not absurd. As TSOs record load–duration curves, the values of these percentages can be considered different for different parts of the day, which generally allows us to talk about peak and valley hours.

However, in the case of wind injection, it is easy to imagine that some wind parks produce one particular percentage of their maximum power and, simultaneously, other parks operate at a very different percentage, especially at sites with complex terrain and when the distances between them are considerable. This was the case, for instance, in Galicia, the autonomous community located in the most northwestern part of Spain, one of the regions where wind energy has experienced the fastest growth over recent years due to its potential [2].

Once this point is reached, the concept of correlation appears, which was soon understood to be the concept that should be used instead of simultaneousness.

Wind speed variability can be classified, in a first stage, by means of two different features: (a) wind speed is different simultaneously at different locations, which results in a correlation between wind speed series; (b) wind speed at the same location is different in time domain, which results in chronological series or autocorrelation processes. The latter is not dealt with here. In this review, we will be focusing on correlation between wind speeds at different locations, as this can be important and useful for reliability studies [3,4].

In general we can say that a first indication of wind correlation at different locations can be expressed as a function of distance, as suggested by Freris and Infield [5].

The problem to solve can be stated thus: is it possible to simulate sets of wind speed values with given correlations and find some features from the point of view of their probability density functions (PDF)? Or, as we are speaking about wind speed distributions, can several Weibull distributions be generated with different shape and scale factors and with the desired correlation matrix?

In this paper, a review is made of different solutions and how they have allowed us to improve the final solution to the problem, which can be helpful when solving subsequent electrical problems such as reliability and probabilistic load flow studies in power networks containing wind parks.

2. Wind speed distributions

As this paper is a review, let us firstly remind ourselves which of the more accepted kinds of statistical distributions are the most widely used for wind speed steady-state simulations.

Meteorological stations measuring wind speed over a period of time (often for at least a year) obtain a good collection of samples, so features such as wind speed mean value and direction are available.

Once the data are collected, the mean wind speed presents a discrete frequency distribution given by the classification of values in an assumed number of intervals.

However, over the years it has been assumed that some continuous distributions can approximate the behaviour of the discrete distribution. Weibull and Rayleigh distributions have been the most generally accepted. The PDF of the Weibull distribution is given in (1), where v_w is the wind speed and C and k are two parameters (scale and shape parameters) that define the shape of the distribution and are related to the mean value, μ, and the standard deviation, σ, by means of the Gamma (Γ) function of Legendre [6], as given in (2) and (3).

$$Pr(v_w) = \begin{cases} 0 & v_w < 0 \\ \dfrac{k}{C}\left(\dfrac{v_w}{C}\right)^{k-1} e^{-(v_w/C)^k} & v_w \geq 0 \end{cases} \tag{1}$$

$$\mu = C\Gamma\left(1+\frac{1}{k}\right) \tag{2}$$

$$\sigma^2 = C^2\left(\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right) \tag{3}$$

As for the derivation of a good Weibull approximation for the distribution of wind speed mean values at a location, the process to obtain it is explained in [7]. In this derivation method [7], C and k parameters have to be sought that fulfill (4), where v_{w0} is a given wind speed and ln is the natural logarithm.

$$\ln(-\ln(Pr(v_w > v_{w0}))) = k \cdot \ln v_{w0} - k \cdot \ln C \tag{4}$$

An alternative way of approximating wind behaviour is the Rayleigh distribution, which is a specific Weibull one, when the shape parameter k equals 2. It generally provides a good approximation and its advantages are, firstly, that it is completely defined by a unique value (the mean value), so it is not necessary to use the Γ function, and secondly, that its use is recommended by international standards [8]. In (5)–(8) the expressions for the Rayleigh distribution can be found.

$$Pr(v_w) = \begin{cases} 0 & v_w < 0 \\ \dfrac{v_w}{a^2} e^{-\frac{1}{2}\left(\frac{v_w}{a}\right)^2} & v_w \geq 0 \end{cases} \tag{5}$$

$$a = \frac{C}{\sqrt{2}} \tag{6}$$

$$\mu = C\frac{\sqrt{\pi}}{2} \tag{7}$$

$$\sigma^2 = C^2\left(1-\frac{\pi}{4}\right) \tag{8}$$

In general, the use of the Weibull approximation can be made when a large amount of previous data exists, whilst the Rayleigh one can be used when only the mean wind speed is known.

3. Simulation of correlated wind speeds according to Pearson parametric correlations

Given two series of wind speeds, corresponding to mean wind speed values every so many seconds at two different locations, different moments from the mean value and the standard deviation can be calculated. One of them is the correlation, which is calculated as a normalized way of giving the variance, i.e., a measure of how both series vary as a set. When the series vary in a very similar way, i.e., when they grow together or decrease together, they will be highly correlated and if the effects are opposite, they are highly correlated in a negative sense.

A definition of the correlation has been given by Pearson, and it is expressed in (9), where ρ is the correlation, σ_{12} is the covariance between series 1 and 2, and σ_1 and σ_2 are the standard deviations of both series (square roots of the variances).

$$\rho = \frac{\sigma_{12}^2}{\sigma_1 \cdot \sigma_2} \tag{9}$$

So, correlation allows us to relate variances (namely, standard deviations) to covariance.

Now we can come back to the questions at the end of the Introduction to this paper and look at the answers to them and how these answers have evolved.

3.1. The method

Let us suppose that we have n wind parks for which we want to have n series of wind speeds, with a given correlation matrix, and also each of them with given shape and scale parameters, C_i and k_i, $1 \le i \le n$.

In [9] two methods were proposed to solve the problem. In this review only what was called method I will be explained, because it is the only one that can be applied with a minimum set of initial values. The method consisted of the following steps:

1. For each wind park, sets of wind speeds are generated by means of the Monte Carlo simulation method, according to the given distribution functions for each of them, i.e., their C and k parameters. The correlation between pairs should be close to 0 as they are stochastically generated.
2. An operation, explained here, is carried out to transform the correlation matrix with values close to 0 to the desired one.

Here is an explanation of both steps and the results obtained in general for them.

Step 1 does not present any difficulties, and it is solved thanks to some commercial software. In fact, the function *wblrnd*() in MATLAB allows us to obtain samples from a Weibull distribution. Assuming the generation of all the Weibull distribution is random, the correlation between all the series should be close to 0.

Step 2 consists of the derivation, for the same samples, of values of correlations close to the desired ones. In method I of [9], the proposal consisted of operating according to a method proposed in [10]. The method was as follows. Let us suppose that we have a vector of uncorrelated variables, $\boldsymbol{z}=(z_1, z_2, \ldots, z_n)^T$, with mean values $\boldsymbol{\mu_z}=(\mu_{z1}, \mu_{z2}, \ldots, \mu_{zn})^T$, and covariance matrix:

$$\Omega_z = \begin{pmatrix} \sigma_{z1}^2 & \sigma_{z12} & \cdots & \sigma_{z1n} \\ \sigma_{z22} & \sigma_{z2}^2 & \cdots & \sigma_{z2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{zn1} & \sigma_{zn2} & \cdots & \sigma_{zn}^2 \end{pmatrix}$$

In this case a new vector $\boldsymbol{y}=(y_1, y_2, \ldots, y_n)^T$ of correlated variables can be obtained, with means $\boldsymbol{\mu_y}=(\mu_{y1}, \mu_{y2}, \ldots, \mu_{yn})^T$ and covariance matrix:

$$\Omega_y = \begin{pmatrix} \sigma_{y1}^2 & \sigma_{y12} & \cdots & \sigma_{y1n} \\ \sigma_{y22} & \sigma_{y2}^2 & \cdots & \sigma_{y2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{yn1} & \sigma_{yn2} & \cdots & \sigma_{yn}^2 \end{pmatrix}$$

if the operation given by (10) is applied.

$$y = L \cdot z + \mu_y \tag{10}$$

In (10), L is a lower triangular matrix fulfilling (11).

$$\Omega_y = L \cdot L^T \tag{11}$$

The matrix L of (10) and (11) is not unique and a good choice for it can be the one obtained with help of Choleski's technique [11,12].

The relationship between the mean values and standard deviations of the sets of variables $\boldsymbol{z}$ and $\boldsymbol{y}$ can be established by means of (12) and (13).

$$E(y) = L \cdot E(z) + \mu_y \tag{12}$$

$$\Omega_y = L \cdot \Omega_z \cdot L^T \tag{13}$$

As one of the objectives of the problem is that $E(y)=\mu_y$, this is obtained when $\boldsymbol{L} \cdot \boldsymbol{E}(\boldsymbol{z})=0$, for which in (12), $\boldsymbol{z}$ can be substituted by $\boldsymbol{z}-\boldsymbol{\mu_z}$, assuming that $E(z)=\mu_z$.

An alternative possible change of variables is to use normalized uncorrelated variables, such as in (14), which allows direct work with correlation matrices instead of covariance matrices, and in this case (13) can be expressed as (14), because $\boldsymbol{\Omega_z}=\boldsymbol{I}$.

$$\Omega_y = LL^T \tag{14}$$

Additionally, when the procedure is applied to normally distributed variables, the resulting vector $\boldsymbol{y}$ is also normally distributed. The numerical results of the method hold not only for multivariate normally distributed variables, but also for all multivariate variables. However, for different types of distributions, the conservation of the kind of distribution cannot be assured, as predicted by the Central Limit Theorem (CLT). The problem of simulating non-normal multivariate distribution functions has been studied in [11], although without a specific solution for the Weibull and Rayleigh cases. In [13] some cases of multivariate distributions with Weibull properties have been studied, and a generalization of the multivariate Rayleigh distribution can be found in [14], but with a great level of mathematical complexity.

When applying the method explained in this section to the simulation of correlated wind speeds, the main problems that appear are the following:

1. Negative wind speeds appear when applying (10).
2. The features of the distributions are lost during the process, although their mean values and correlations are kept.

For a better understanding, let us express (10) such as in (15).

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \tag{15}$$

In (10), i.e., in (15), the values of $\boldsymbol{z}$ are randomly obtained, so a number of samples has to be generated to simulate the complete statistical process. For example, each z_i can represent the value of the wind speed at one wind park, i, at a given time. So, a number of samples must be generated for that z_i if a Monte Carlo simulation has to be carried out.

The values y_i of $\boldsymbol{y}$ are sequentially calculated, first y_1, then y_2, and so on, until y_n. By doing this to all the samples, the expected result is a set of values for the variables y_i with the given correlation (covariance) matrix.

Negative values can appear during the process, depending on the values of the matrix $\boldsymbol{L}$. This is a problem because negative wind speed values do not exist.

And when the process finishes, experience shows that the feature of the non Normal distribution is progressively lost as the subindex number of the variable grows. So, if the initial distributions were Weibull distributions, then, y_1 of the correlated variables will be a Weibull distribution, but y_n will be far from being one.

In [9] some simplifications were assumed to avoid the problems that appeared.

As the number of sets with negative values of wind speeds was generally between 5 and 10%, the solution was not to take these series into account. So, only the other 90–95% of the series with simulated correlated wind speeds were taken into account. This procedure adds some error to the global solution, but it is assumed as part of the process.

On the other hand, instead of Weibull distributions, and according to recommendations given in [8], normalized Rayleigh distributions were assumed to represent all the wind speed distributions, which at least meant that all the initial distributions had similar features. However, they degenerate like Weibull distributions.

3.2. Error measurement

In the previous subsection we have mentioned the errors made when applying the proposed method to the simulation of correlated wind speeds. There are two kinds of errors:

1. Errors in the mean value or in the correlation/covariance matrix
2. Errors due to the loss of features in the distributions.

The first kind can be assessed numerically, by comparison between the desired mean values with the obtained mean values or between the desired correlation matrix and the obtained one. Although the initial sets contained in z present mean values very close to those desired, the operation given in (10) can introduce changes in these values. And if the decision is to cancel all the simultaneous wind speed series where at least one of the values is negative, then the error will be greater.

However, as those errors can be numerically assessed, we are going to focus here on the second kind of errors, those due to loss of features in the distributions. Due to the CLT, as we have mentioned, and due to the operational procedure, explained with help of (13), we can say that when the number of distribution functions grows (tends to n), this distribution tends to a Normal one (which explains the appearance of negative values). But if the initial distribution was a Rayleigh or Weibull one, then it is obvious that there is a difference between the desired distribution and the obtained one. In order to assess the error due to this, a statistical fitness method can be used, for example based on the χ^2 distribution. With commercial software such as MATLAB and Weibull distribution functions, this can be done directly by means of the function *wblfit*().

3.3. Example

As an example, let us suppose that we want to simulate 10,000 values of wind speed for three locations where the C and k values for the Weibull distributions are $C_1=8$, $k_1=2$, $C_2=7$, $k_2=2.4$ and $C_3=6$, $k_3=1.7$, the mean values are 7.1207, 6.2380 and 5.3226 respectively, and the desired correlation matrix is the following one:

$$\Omega_y = \begin{pmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.5 \\ 0.6 & 0.5 & 1 \end{pmatrix}$$

After carrying out the process, the results are three distributions with mean values exactly the same as the desired ones, and correlation matrix:

$$\Omega_y = \begin{pmatrix} 1 & 0.7920 & 0.6487 \\ 0.7920 & 1 & 0.5697 \\ 0.6487 & 0.5697 & 1 \end{pmatrix}$$

However, from the 10,000 series of three simulated wind speeds, all them positive, at least one negative value appears in 460 of them. If we reject these series, which are less than 5% of the total, then we will have three series without negative values. In that case, the mean values change to 7.3555, 6.4519, and 5.5829, and the parameters obtained in the simulation for the distributions are $C_1=8.3099$, $k_1=2.1530$, $C_2=7.2632$, $k_2=2.1263$, $C_3=6.2395$ and $k_3=1.7311$ respectively. The correlation matrix in this case is:

$$\Omega_y = \begin{pmatrix} 1 & 0.7790 & 0.6164 \\ 0.7790 & 1 & 0.5348 \\ 0.6164 & 0.5348 & 1 \end{pmatrix}$$

We can summarize the results of the process in Table 1, where subindex s means specified (desired) and c means calculated. WP means wind park and the subindex of each of one is its number. In the case of mean values, μ_s is the mean value of the 10,000 samples of the initial uncorrelated set, and μ_c the mean value of the samples of the final correlated set, generally less than 10,000 due to elimination of sets with negative values. μ_Γ is the value obtained from (2) by means of the Gamma function.

Table 1
Proposed and calculated values in the simulation.

	C_s	C_c	k_s	k_c	μ_Γ	μ_s	μ_c
WP_1	8	8.3099	2	2.1530	7.0898	7.1207	7.3555
WP_2	7	7.2632	2.4	2.1263	6.2054	6.2380	6.4519
WP_3	6	6.2395	1.7	1.7311	5.3535	5.3226	5.5829

4. An improvement of the method: the use of evolutionary algorithms

A different point of view in the simulation of correlated wind speeds is given in [15], and upgraded in [16]. In both works evolutionary methods are applied and the results can be pointed out.

4.1. Description of the method

The method proposed in [15] consists of generating series of wind speeds, one for each location, fulfilling the distribution properties, and then to re-arrange them, keeping the distribution features for the corresponding locations, but obtaining the correlation matrix requested. So we have to make the observation that the initial generated values will not change, but only their positions in the vectors will change during the process.

In order to carry it out, evolutionary techniques are used, such as Genetic Algorithms (GA) and Local Search (LS). GA has been explained in [17] and LS in [18]. It can be said, generally speaking, that the evolutionary techniques operate by minimizing some kind of error function, starting from a possible solution, and modifying it until another condition or conditions are reached, whilst maintaining others from the possible solutions. This process is performed repeatedly if the error function reduces its value, until the goal is reached.

In the case of the simulation of wind speed series, the condition maintained throughout the process is the conservation of the PDF of the wind speed series for every single location, based on the fact that the change of position of the elements in their sets does not modify the PDF. The conditions that have to be fulfilled are the correlations between each pair of wind speed series, defined in the correlation matrix. The error measure functions, defined as the proximity to the desired results, are obtained through the maximum norm of the differences between the elements of the correlation matrix, obtained from the possible solution, and the desired one. Note that this kind of error measure could be used in the method explained previously, although it would not serve to improve it, as it does in the evolutionary techniques. So, for the possible solution k, the error function is defined in (16), where ρ_{ij} is the desired value of the correlation between locations i and j and ρ_{ij}^k is the value of this correlation in the k-th iteration of the process.

$$d^k = \max\left\{\rho_{ij} - \rho_{ij}^k\right\}, \quad 1 \le i, j \le n \tag{16}$$

The evolutionary techniques in [15] are very common and complementary. LS is based on the premise that a possible solution has better ones around it, so the first one is perturbed, maintaining certain conditions. If the solution obtained is better, it is accepted, if not, it is discarded, and the original is kept for the next iteration. GA consists of crossing two possible solutions, speeding the evolution of the solution if a better solution is obtained than the ones used. Both techniques also consider accepting some kind of worse solution, in order to avoid the possibility of reaching a local minimum.

When applied, the LS consists of swapping pairs of wind speed values in the same corresponding location, in order to improve the correlation matrix of the whole group. GA consists of generating possible solutions of wind speed series combining series from two other locations, seeking the same improvement.

Perhaps the great disadvantage of using evolutionary techniques for the solution of such a problem is the fact that they need a long computing time. Some possible ways to reduce this are the options to initialize the method described in next section.

4.2. Options to initialize the method

The key in any evolutionary or iterative process is good initialization, in order to avoid wasting time avoiding local minimums of the error function. This happens in the method described in [15], where a lot of computation time is consumed, especially as so many locations are considered.

Two types of initialization are described in [16] in order to reduce the time consumed in [15]. They will be called fixed and optimum initialization methods.

4.2.1. Fixed initialization

In order to generate the wind speed samples of two locations, when there is a correlation greater than $\rho_{ij} > 0$ between them, a possible interpretation of the value of this correlation is the percentage of wind speed data with full correlation, i.e. $\rho_{ij} = 1$, considering that the rest of the values have a correlation $\rho_{ij} = 0$. The data with full correlation can be obtained from the same Uniform distribution for both locations, inverting each Cumulative Density Function (CDF) and utilizing their respective parameters, (C_i, k_i) and (C_j, k_j). The data with correlation 0 are generated independently for each location.

Therefore, the initial possible solution, which is the input to the method described, can have an initial correlation matrix where all of its non-diagonal elements are equal to a fixed value (for example, 0.5). Moreover, as the method described utilizes two possible solutions, two initializations of this type with different values are suggested.

Obviously, the best initialization depends on the values of the correlation matrix, but combined values of 0.6 and 0.4 or 0.7 and 0.3 are recommended. Generally speaking, it can be said that messing up is easier than arranging, so higher values reduce the computation time significantly.

4.2.2. Optimum initialization

In order to approximate the initial possible solution provided as an input of the method described, the fixed initialization can be improved. Taking the whole group of locations and its different correlation values by pairs, following what has been explained, they can be understood as percentages of full correlation values.

First the whole group of n locations is considered, and the minimum correlation value is assumed to be the percentage of values with full correlation for all of them, generating this quantity of values, related, for all locations. Then, all the groups of $n-1$ locations are considered, obtaining the percentage of values totally correlated, where the percentage considered for n have to be deduced, and this quantity of values are generated for each corresponding group of $n-1$ locations, whilst for the other one, randomly generated values are provided. The process can be repeated for groups of $n-2$, $n-3$, ... and so on, but the smaller the groups, the more time it takes to obtain the initialization. It is recommended that this process just considers groups at n, $n-1$ and $n-2$ locations.

Finally, it can be said that the method described can provide a solution that is as approximate as required. The only circumstance to consider is the computation time, which depends on the number of locations considered, the initialization utilized, and, of course, the speed of the computer, but future improvements in computer technology will minimize the need for this consideration.

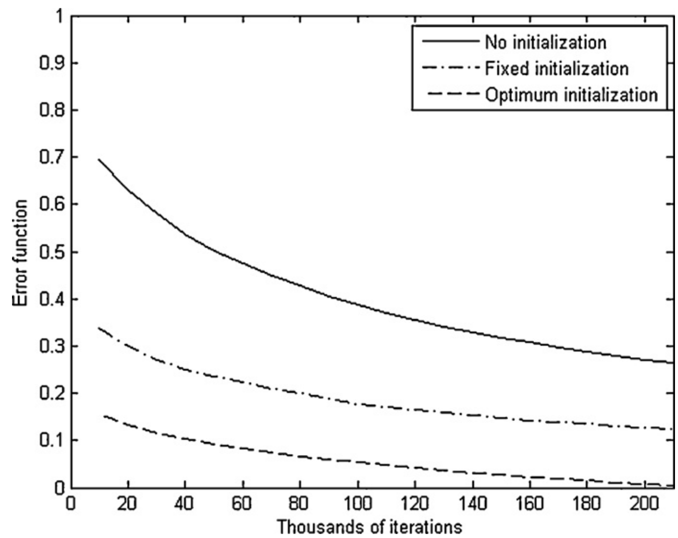


Fig. 1. Evolution of the algorithm considering different types of initialization.

4.3. Initialization evaluation

In order to show the evolution of the algorithm considering its initialization, Fig. 1 is provided, where the evolution of the error function is plotted against the number of iterations. In comparison to no initialization, the fixed initialization starts the iteration process closer to the objective, reducing the computation time. The same happens for the optimum initialization, but in this case, the process starts even closer to the objective.

5. Simulation of correlated wind speeds according to Spearman rank correlations

The proposal here is to change the definition of correlation being used. So far we have talked about parametric Pearson's correlation. However, if a different correlation parameter is used, i.e. the non-parametric Spearman rank correlation, the problem of derivation of correlated wind speeds can be better solved. Spearman rank correlation is not calculated as a function of the values, but on the basis of the ranking of these values in the series.

5.1. The method

The algorithm was presented in [19] and is based on an interesting statistical work [20], by Iman and Conover (IC). The results of the IC work demonstrate that correlated distributions can be simulated by means of the Monte Carlo technique in a more accurate way than that described in [13].

Spearman rank correlation can be expressed as in (17).

$$r = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} \tag{17}$$

where the samples of both series have been previously ranked, and d_i are the differences in rank between samples of both series that occur simultaneously, and n is the number of samples.

As a summary, the method consists of the following process.

Firstly, the Spearman rank correlation between all pairs of wind speed series is calculated according to (13). A Spearman rank corre-

lation matrix is obtained in the same way as a Pearson's correlation matrix was obtained when using method I previously explained.

Secondly, the Monte Carlo technique is used to generate the wind speed distributions corresponding to the different locations, independently.

Finally, the IC method is used as described in [20] to obtain sets of simultaneous wind speeds with the desired correlations. This method is as follows:

1. A number of independent distribution functions coinciding with the number of distributions of the original problem is generated (Weibull or Rayleigh). Also, the number of samples must coincide.
2. The operation $\boldsymbol{y} = \boldsymbol{L} \cdot \boldsymbol{z}$, where $\boldsymbol{z}$ represents independent Uniform distributions, $\boldsymbol{y}$ is the vector of correlated Uniform distributions and $\boldsymbol{L}$ is a lower triangular matrix obtained from the correlation matrix (Spearman indices) by means of the Choleski technique.
3. Some order of the elements is analyzed in each vector. For example, if they are sorted by their values, a classification is established according to their ranks in the vectors.
4. After this, the same order is analyzed in the original distributions, and they are re-ordered according to a ranking similar to that obtained in step 3.

The results of step 2 can be values out of the range of the original Uniform distributions. This is not important, as all that will happen is that similar ranks will be used in the Weibull or Rayleigh distributions, which, when applied to the generated independent distributions, will only reorder their values in the vectors.

As a result, correlated distributions are obtained according to the rank correlation matrix desired and also keeping all their features, i.e., if they were Weibull distributions, they will remain Weibull distributions. This is due to the fact that all their values are respected, and only their rank values (their positions in the series) are changed.

For example, if a first set has the following values: $\boldsymbol{z_1} = [1,3,6,7,8]$, and the second one $\boldsymbol{z_2} = [4,2,6,1,3]$, and if we reorder this second one and we rewrite it as $\boldsymbol{y_2} = [6,1,3,2,4]$, it is obvious that the CDF and, of course, the mean value and standard deviation corresponding to $\boldsymbol{y_2}$ have not changed with respect to $\boldsymbol{z_2}$, but the correlation between $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ is different from the correlation between $\boldsymbol{z_1}$ and $\boldsymbol{y_2}$.

5.2. Example

In order to compare the results with those obtained with the first method presented, they are given for its application to the example with the same wind parks, with a small difference: the change of the correlation definition. According to this, the values that were used in the first method as Pearson parametric correlations are used here as Spearman rank correlations.

The resulting Spearman rank correlation matrix is as follows:

$$\Omega_{yS} = \begin{pmatrix} 1 & 0.6871 & 0.5807 \\ 0.6871 & 1 & 0.4741 \\ 0.5807 & 0.4741 & 1 \end{pmatrix}$$

Although it was not the goal of this method, we can add that the Pearson parametric correlation matrix obtained as a result was the following:

$$\Omega_{yP} = \begin{pmatrix} 1 & 0.6810 & 0.5840 \\ 0.6810 & 1 & 0.4935 \\ 0.5840 & 0.4935 & 1 \end{pmatrix}$$

What can be commented about this Pearson correlation matrix is that it does not seem to be a worst approximation to the initial Pearson correlation matrix obtained by means of the first method presented. So, even if only Pearson parametric correlations are known, perhaps the use of this second method can be advantageous, by assuming they are Spearman rank correlation values.

Table 2
Proposed and calculated values in the simulation.

	C_s	C_c	k_s	k_c	μ_Γ	μ_s	μ_c
WP_1	8	7.9878	2	1.9795	7.0898	7.0825	7.0825
WP_2	7	6.9981	2.4	2.4124	6.2054	6.2032	6.2032
WP_3	6	5.9960	1.7	1.7159	5.3535	5.3464	5.3464

And finally, the results obtained for the other parameters are given in Table 2. We can observe that the parameters of the final distributions are close to the samples of the proposed one, and the mean values of the initial and final distributions coincide. Furthermore, if the initial sets contain 10,000 samples, the final ones also do, as there is no elimination of values, as all of them are positive numbers.

6. Other related works

Part of the work that has not been included in this research is the use of Markov chains to obtain correlated wind speeds. In a way, method II presented in [9] can be said to have some relationship with them, although it is not based on them, and was not commented here.

Alternatively, for those who have simultaneous data of wind speeds, the most obvious way to simulate is by repeating these values as if they were randomly generated. It is the simplest and quickest way. The methods proposed and commented here are envisaged for cases when the initial data are mean wind speeds and correlations.

One of the main applications for the techniques presented in this review, is to generate wind power series of aggregated wind turbines in order to obtain wind power distributions of wind parks. Some results of this have been presented in [21]. In [22] a study of the wind power distributions and their applications has been presented.

7. Conclusions

After long experience simulating series of wind speeds at different locations, the following conclusions can be made from a comparison of results obtained by the different means presented in this review.

The main conclusion is that when it is not mandatory to use parametric correlations, but rank correlations can be used, then the method based on the Spearman rank correlation definition is perhaps best for simulation, as it is very quick and allows us to simulate whilst keeping the correlation matrix and the features of the distributions.

When parametric correlations have to be used, then the methods based on genetic algorithms offer good results, and they can be optimized in order to be run in a short time.

Finally, there should be no need to use the first methods employed, and they were described because they formed the origin of this research.

Acknowledgements

The financial support given by the autonomous government Xunta de Galicia, under the contract INCITE 08REM009303PR is gratefully acknowledged by the authors. This work was also partially supported under the project S/WE/4/08 at the Bialystok University of Technology.

References

[1] The European Wind Energy Association: Energy Report, www.ewea.org; 2009.
[2] Troen I, Petersen EL. European Wind Atlas. Denmark: Risø National Laboratory; 1989.
[3] Wen J, Zheng Y, Donhang F. A review on reliability assessment of wind power. Renewable and Sustainable Energy Reviews 2009;13:2485–94.
[4] Vallée F, Lobry J, Deblecker: O. Impact of the wind geographical correlation level for reliability studies. IEEE Transactions on Power Systems 2007;22 (November (No. 4)).
[5] Freris L, Infield D. Renewable energy in power systems. John Wiley and Sons Ltd.; 2008.
[6] Bronshtein I, Semendyayev K. Handbook of mathematics. Mir 1973.
[7] Freris LL. Wind energy conversion systems. Prentice Hall; 1990.
[8] IEC 62400-1Wind Turbine Generator Systems. Part 1: Safety Requirements.
[9] Feijóo A, Cidrás J, Dornelas JLG. Wind speed simulation in wind farms for steady-state security assessment of electrical power systems. IEEE Transactions on Energy Conversion 1999;14(December (No. 4)):1582–8.
[10] Kleijnen J, van Groendall W. Simlation. A statistical perspective. Wiley and Sons; 1992.
[11] Johnson ME. Multivariate statistical simulation. Wiley and Sons; 1987.
[12] Naylor TH, Balintfy JL, Burdick DS, Kong C. Computer simulation techniques. NY: Wiley; 1966.
[13] Lee L. Multivariate distributions having Weibull properties. Journal of Multivariate Analysis 1979;9:267–77.
[14] Jensen DR. A generalization of the multivariate Rayleigh distribution. Sankhya: The Indian Journal of Statistics 1970;A(No. 4):193–208.
[15] Villanueva D, Feijóo: A. A genetic algorithm for the simulation of correlated wind speeds. International Journal of Electrical Energy Systems 2009;2:107–12.
[16] Villanueva D, Pazos JL, Feijóo A. A method for improving the simulation of correlated wind speed series, accepted for publication in the Scientific Proceedings of EWEC 2011.
[17] Gen M, Cheng: R. Genetic algorithms and engineering. New York: John Wiley and Sons; 2000.
[18] Hoos HH, Stützle T. Stochastic local search. Foundations and applications. Morgan Kaufmann/Elsevier 2004.
[19] Feijóo A, Sobolewski R. Simulation of correlated wind speeds. International Journal of Electrical Energy Systems 2009;1(July–December (n2 2)): 99–106.
[20] Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. Communications on Statistics Simulation and Computation 1982;11:311–32.
[21] Villanueva D, Feijóo A, Pazos JL. Correlation between power generated by wind turbines from different locations. Proceedings of EWEC 2010.
[22] Villanueva D, Feijóo A. Wind power distributions: a review of their applications". Renewable and Sustainable Energy Reviews 2010;14:1490–5.

Andrés Feijóo obtained his Ph.D. degree in Electrical Engineering from the *Departamento de Enxeñería Eléctrica, Universidade de Vigo*, Galicia, Spain, in 1998. He continues in this department and is interested in renewables and, in particular, the influence of wind energy in the electrical network, and steady-state and dynamic modelling and simulation of electrical machines for wind parks.

Daniel Villanueva has an MSc degree in Electrical Technology and Electronics Engineering from the *Universidade de Vigo*, Galicia, Spain, and he is currently working on the impact of wind energy in power systems, including simulation of wind speeds, analysis of wind power, and probabilistic analysis of power systems including wind power generation.

José Luis Pazos has an MSc degree in Mining Engineering, specializing in Energy. He is now a Ph.D. student at the *Departamento de Enxeñería Eléctrica, Universidade de Vigo*, Galicia, Spain, and he is developing his research into the impact of wind energy in electrical power systems.

Robert Sobolewski received his Ph.D. degree in Electrical Engineering from Bialystok University of Technology in 2002. At present he works with the same University, Department of Power Engineering. His current research interests include: reliability and risk analysis in power systems with the dispersed generation integrated and reliability of protective relays.